Land Information System - An Interoperable Framework for High Resolution Land Surface Modeling ⋆

S. V. Kumar [a], C. D. Peters-Lidard [b], Y. Tian [a], P. R. Houser [b], J. Geiger [b], S. Olden [b], L. Lighty [b], J. L. Eastman [a], B. Doty [c], P. Dirmeyer [c], J. Adams [c], K. Mitchell [d], E. F. Wood [e], J. Sheffield [e]

[a] *University of Maryland at Baltimore County, Baltimore, MD 21250*

[b] *NASA Goddard Space Flight Center, Greenbelt, MD, 20771*

[c] *Center for Ocean-Land-Atmosphere Studies, Calverton, MD 20705*

[d] *NCEP Environmental Modeling Center, NOAA/NWS, Camp Springs, MD*

[e] *Princeton University, Princeton, NJ*

Abstract

Knowledge of land surface water, energy, and carbon conditions are of critical importance due to their impact on many real world applications such as agricultural production, water resource management, and flood, weather, and climate prediction. Land Information System (LIS) is a software framework that integrates the use of satellite and ground-based observational data along with advanced land surface models and computing tools to accurately characterize land surface states and fluxes. LIS also employs the use of scalable, high performance computing and data management technologies to deal with the computational challenges of high resolution land surface modeling. To make the LIS products transparently available to the end users, LIS includes a number of highly interactive visualization components as well. The LIS components are designed using object oriented principles, with flexible, adaptable interfaces and modular structures for rapid prototyping and development. In addition, the interoperable features in LIS enable the definition, intercomparison, and validation of land surface modeling standards and the reuse of high quality land surface modeling and computing system.

1 Introduction

Land surface water, energy, and carbon conditions have profound influences on the overall behavior of the climate systems. A better understanding of these conditions helps in the improved use of natural resources, prevention of adverse impacts, and our adaptation to climate change. Researchers have been involved in integrating land surface simulation, observation, and analysis methods to accurately determine land surface energy and moisture states. Examples of such systems include the 1/8 degree North American Land Data Assimila-

⋆ Corresponding author: S. V. Kumar. Tel. 001-301-286-8663.

Fax: 001-301-286-8624, *email: sujay@hsb.gsfc.nasa.gov*

tion System (NLDAS) [Mitchell et al., 2004] and the 1/4 degree Global Land Data Assimilation System (GLDAS) [Rodell et al., 2004]. Computational limitations in hardware and software have impeded the development and application of such systems at higher spatial resolutions. LIS is a software system that takes advantage of the technological improvements in computing and environmental monitoring tools to enable a global high resolution (down to 1km) land modeling system. The high resolution modeling capabilities enables LIS to directly ingest the vast array of high resolution observations such as those available from the next generation NASA earth science instruments (Earth Observing System (EOS) Terra and Aqua). The ability to operate at the same fine spatial scales of the atmospheric boundary layer and cloud models also helps in improving water and energy cycle prediction capabilities. In addition to providing a land surface modeling infrastructure, the portable, interoperable design of LIS enables it to be a valuable research tool for land surface researchers and other interdisciplinary scientists.

The land surface modeling infrastructure in LIS consists of several land surface models (LSMs) run typically in an uncoupled manner, using a combination of observationally-based precipitation and radiation with downscaled model-based meteorological inputs and many surface parameters. Simulation of land surface processes using these models at high spatial resolution is computationally demanding due to the large number of simulation runs required and the relatively high data density of the LSMs. LIS makes use of the state-of-the-art scalable high performance computing technologies to overcome these challenges. To provide efficient management, storage, and high throughput data access in simulations, LIS also employs a number of generic tools to manage the input and output data. To enable the effective use of the system to end users, LIS also provides intuitive web-based interfaces to LIS data and other resources.

Many existing earth science applications, though highly scalable and computationally capable, lack the ability to interoperate with other earth system applications. As a result, the cost of adding new functionalities and adapting the existing systems to function with other applications may be prohibitively high. LIS attempts to achieve code interoperability by applying advanced software engineering concepts in its design. The system is designed as an object oriented framework that can be shared and reused by scientists and practitioners in the land surface modeling community. LIS provides the use of a complete, usable, and integrated set of high level tools that can be applied without the necessary knowledge of underlying computer hardware or software. The use of object oriented principles help in designing LIS to be flexible and extensible, enabling rapid prototyping of new applications into LIS.

In addition to providing an infrastructure to support land surface research and applications activities, LIS has also adopted other earth system modeling standards and conventions, such as the Earth System Modeling Framework (ESMF) [Hill et al., 2004] and Assistance for Land Modeling Activities

(ALMA) [ALMA, 2002]. ESMF is a system that provides a flexible software infrastructure to foster interoperability, portability, and code reuse in climate, numerical weather prediction, data assimilation, and other earth science applications. ALMA is a land-atmosphere coupling standard that is being developed by the broad land-atmosphere research community. By conforming to the standards laid out by ESMF and ALMA, LIS provides capabilities to interact with other earth system models.

The following sections describe the land modeling and computing tools in LIS, the interoperable features and adoption of earth system modeling standards, and the application of LIS in modeling land surface processes.

2 Land Surface Modeling in LIS

Land surface modeling seeks to predict the terrestrial water, energy, and biogeochemical processes by solving the governing equations of the soil-vegetation-snowpack medium. The land surface and atmosphere are coupled to each other over a variety of time scales through the exchanges of water, energy, and carbon. An accurate representation of land surface processes is critical for improving models of the boundary layer and land-atmosphere coupling at all spatial and temporal scales and over heterogeneous domains. Long term descriptions of land use and fluxes also enable the accurate assessments of climate characteristics. In addition to the impact on the atmosphere, predicting land surface processes is also critical for many real-world applications such as ecosystem modeling, agricultural forecasting, mobility assessment, and water resources prediction and management.

A schematic representation of land surface modeling in LIS is shown in Figure 1. LSMs typically require three types of inputs: 1) Initial conditions, which describe the initial state of the land surface; 2) Boundary conditions, which describe both the upper (atmospheric) fluxes or states also known as “forcings” and the lower (soil) fluxes or states; and 3) Parameters, which are functions of soil, vegetation, topography, and other surface properties. Using these inputs, LSMs solve the governing equations of the soil-vegetation-snowpack medium and predict surface fluxes (sensible, latent, ground heat, runoff, evaporation) and soil states (moisture, temperature, snow), providing a realistic representation of the transfer of mass, energy, and momentum between a vegetated surface and the atmosphere [Sellers et al., 1986]. The model results are aggregated to various temporal and spatial scales to assess water and energy balances. The results can also be compared with in-situ observations if available.

3 Components of LIS

The LIS software system consists of a number of components: (1) LIS core: the core software that integrates the use of LSMs, high performance computing, use of various sources of data, and the domains of execution; (2) A number of community LSMs; (3) Data servers to provide a common interface to heterogeneous data and handle access requests and (4) Visualization tools to provide interactive access to the LIS products. Various software components of LIS are shown in Figure 2.

LIS core, the central part of the system, is primarily an infrastructure that operates multiple one-dimensional LSMs providing it the appropriate inputs. These models are typically run in an uncoupled manner, where the boundary conditions for the atmosphere are provided either from meteorological forecast model outputs or from various satellite and ground-based observational systems. The input data, including the initial conditions and model parameters derived from topography, vegetation, and soil coverage describing the land surface states are processed and supplied to the LSMs. The models in turn produce optimal output fields of land surface states and fluxes.

The LSMs in LIS use and produce numerous data for analysis and modeling purposes in different data formats and resolutions. LIS provides a number of generic data management utilities to ensure a seamless, efficient access and use of data. The heterogeneity of diastases are encapsulated by the use of data servers based on Grid Analysis and Display System (GrADS [Doty and Kinter, 1993])-DODS (Distributed Oceanographic Data System) (GDS [Wielgosz et al., 2001]). The client-server model of data serving provided by the GDS server and GrADS clients allows distributed data sets in various formats to be accessed dynamically and transparently. LIS also provides tools to retrieve input data from various sources, interpolate, reproject, and subset them to the required domain and resolution. The LIS datasets include satellite and remote-sensing land surface data, and products such as temperature, Normalized Difference Vegetation Index (DVI) greenness, Leaf Area Index (LAI), surface albedo, and emissivity from Advanced Very High Resolution Radiometer (AVHRR), Geostationary Operational Environmental Satellite (GOES), Moderate Resolution Imaging Spectroradiometer (MODIS) and various airborne sensors. The ability to spatially resolve resolutions down to 1km and finer enables the direct use of the high resolution data produced by satellite technologies such as MODIS instrument on Terra and Aqua satellites in LIS. Figure 3 shows a sample of the 1km global MODIS leaf area index product used in LIS.

Land surface modeling at high spatial resolutions such as 1km presents considerable computational challenges. Typically, the land surface is modeled by dividing it into two -dimensional regions or gridcells (for example, cells of size lkm $\times$ 1km globally would lead to approximately 5×10^8 gridcells). Assuming approximately 15 milliseconds for each day of land surface model execution on a single gridcell with 15 minute timesteps, it can be estimated that to conduct a day's simulation at 1km on a single processor would require approximately 3 months. Further, as the number of grid points increases with resolution, the memory and disk storage requirements also increase significantly. For the global 1km simulation, it can estimated that the memory and disk storage requirements scale to the order of Terabytes.

Due to the significant computational requirements, the use of scalable computing technologies is critically important for LIS, especially at high spatial resolutions. Land surface processes have rather weak horizontal coupling on short time and large space scales. LIS exploits this inherent parallelism to achieve highly efficient scaling across massively parallel computational resources. To adequately address the computational requirements at different resolutions and domains, LIS provides a number of high performance operating modes. For resolutions and domains where the available memory is not a limiting constraint, LIS employs a simple master slave paradigm, with a master processor performing the initializations and domain decompositions. Subsequently, the slave nodes perform computations on the decomposed domain. The temporal synchronizations and output aggregations are conducted by the master processor during a simulation at specified intervals. However, this mode becomes intractable for large domains at high spatial resolution. For such cases, LIS makes use of the GDS data servers to handle the I/O. GDS provides capabilities for a client to dynamically retrieve subsets of data on a global domain. The domain decomposition is done as before by a master processor, and the initializations and subsequent computations are performed by the slave nodes. These slave nodes request the required subset of data from the GDS data server as the computations proceed. The data is retrieved from the server by each slave node using a GrADS client. The GDS server performs output aggregations subsequent to the completion of computations by the slave nodes. An illustration of the roles of the master, slave, and the GDS servers in the two modes of operation is shown in Figure 4.

The goal of the user interface components in LIS is to allow the interactive, flexible use of the LIS products to the end users. The visualization capabilities in LIS are built based on a multi-tier client-server system architecture. GDS data servers are employed to handle various types of client requests, such as web-based and DODS-based. On the client side, the user can use different types of client programs as the front end: a web browser or a DODS client program. LIS uses the capabilities of the GDS server to handle DODS client requests, and a web server to handle others. When a DODS client is used, the user can also perform data manipulations such as subsetting and dynamic generation of images. LIS also provides an alternate method to visualize the LIS data using the Live Access Server (LAS [Hankin et al., 2001]). LAS is a highly configurable Web server designed to provide flexible access to geo-referenced scientific data. The LAS server's abilities allow users to search a data catalog, visualize data interactively, request subsets, view metadata, and a multitude of other functions.

4 Interoperability in LIS

LIS is designed embodying the software-development practices to encourage the reuse and community sharing of scientific modeling algorithms. LIS is a framework to combine land surface models, relevant data, and computing tools and resources. These components are designed as several functional abstractions using the flexible paradigms of object-oriented programming to facilitate reuse and development of future extensions. The interoperable features in LIS also include the reuse and participation with other earth system modeling groups. The following sections describe these aspects of LIS design.

4.1 Interoperable Design Features in LIS

As defined in the software engineering literature [Nowack, 1997], an object-oriented framework represents a software system designed for a family of problems and provides a reusable design for applications within that domain. The reusable, "semi-complete" nature of object oriented frameworks makes it easier to build correct, portable, efficient, and inexpensive applications. An object oriented framework normally provides a number of points of flexibility in the design called "hot spots" [Pree, 1995]. Hot spots are abstract methods that must be implemented in order to use the framework for a specific application. The parts of a framework that cannot be altered are called the kernel or frozen spots. The use of hot spots provides implicit reuse of high quality proven software. By incorporating these principles into design, LIS provides an "off-the-shelf" framework for land surface modeling applications.

The LIS software is primarily written in Fortran 90 programming language. Fortran 90 provides a number of features that are useful for object oriented style of programming such as derived types, modules, and generic interfaces, but lacks the support for object oriented properties such as inheritance and run-time polymorphism. However, it is possible to emulate these properties in software [Decyk et al., 1997], enabling an object-oriented programming style in Fortran 90. The compile-time polymorphism in LIS is simulated by the use of virtual function tables. C language allows the capabilities to store functions; a Fortran 90 program can interface with C to store Fortran 90 functions to be invoked at runtime. By combining the features of both these languages, LIS uses a complete set of operations with function pointers.

The overall LIS design incorporates many object oriented principles, such as encapsulation of data and control, inheritance, and compile-time polymorphism. Similar to the "semi-complete" nature of the object oriented frameworks, LIS design provides common functionalities for land surface modeling, leaving the variable functionalities to be filled in by the user. The number of variable functionalities in LIS include: interfaces to facilitate the incorporation of (1) domains, (2) LSMs, (3) land surface parameters, and (4) meteorological input schemes. The LIS software architecture follows a layered pattern as shown in Figure 5. The top layer handles operations related to the overall program control and a number of generic tools. Figure 6 shows the different functions performed by the LIS core. These include operations related to the overall control, runtime statistics, interlanguage support, error logging and dynamic memory management functions. Routines to manage domain decomposition, load balancing, fault tolerance, etc. are also encapsulated as generic routines in the high performance computing and communications (HPCC) component. The time management tools in LIS are built based on the ESMF time management utility, which provides useful functions for time and data calculations and higher level functions to control model timestepping and alarms. Another tool implemented in the LIS core structure is the generic I/O tool, which provides capabilities to read the input data locally or through a GDS-based data server. The I/O tools also provide support for distributed data output and multiple formats. Other miscellaneous tools incorporated in the top layer include methods to perform spatial and temporal interpolation, reprojection, domain subsetting etc. The abstractions providing representations for the behavior of LSMs, domains, and data are also incorporated in the top layer.

The middle layer provides a number of functional abstractions to represent the variable functionalities in LIS. The "plugin" interfaces, `domain-plugin`, `lsm-plugin`, `forcing-plugin`, and `param-plugin`, contain hot spots or extensible interfaces for incorporating new domains, LSMs, meteorological forcing schemes, and parameter data, respectively. The lowest layer contains the customized, user-defined implementations of each component. For example, the Variable Infiltration Capacity (VIC) is a land surface model that is implemented in LIS by extending the `lsm-plugin` interfaces. Similarly, other user-defined components use corresponding extensible interfaces. These components include models and data schemes from National Center for Atmospheric Research (NCAR), National Center for Environmental Prediction (NCEP), University of Washington (UW), Princeton University (PU), NASA Goddard Space Flight Center (GSFC), Air Force Weather Agency (AFWA), National Weather Service (NWS) Climate Prediction Center (CPC), University of Maryland (UM), United States Geological Survey (USGS), National Oceanic and Atmospheric Administration (NOAA), Natural Resources Conservation Service (NRCS), and Boston University (BU). Table 1 lists a summary of the growing list of LSMs and data schemes implemented in LIS.

The adaptable interfaces in LIS described above enable the reuse of the broad set of data, high performance computing, data management, visualization tools, and the land modeling infrastructure in LIS. Further, the LIS framework allows researchers to perform intercomparisons of model output and sensitivity experiments of different LSMs, meteorological scheme, and input on various domains.

4.2 Adopted interoperable features in LIS

To interoperate with other scientific modeling communities, LIS has adopted a number of modeling standards and other frameworks in its design. ALMA is a flexible data exchange convention that LIS adopts to facilitate the exchange of forcing data for LSMs and the results produced by them. The output data variables and formats, and the variables passed between LIS and the land models follow the ALMA specification. The implementation of ALMA conventions allows LIS to exchange data with other land modeling systems that are also ALMA compliant. Further, ALMA compliance enables LIS to be used for intercomparisons of LSMs.

Another interoperable component that LIS uses is the ESMF. ESMF is a framework that provides a structured collection of building blocks that can be customized to develop model components primarily for Earth Science applications. ESMF can be broadly viewed as consisting of an infrastructure of utilities and data structures for building model components and a superstructure for coupling and running them.

ESMF provides a utility layer that presents a uniform interface for common system functions such as time manager, basic communications, error handler, diagnostics, etc. LIS uses a number of ESMF utility tools that has enabled the reuse of software as well as ease of development. The ESMF superstructure provides mechanisms to address physical consistency between data that are represented differently inside the individual components, or partitioned differently on a parallel computer. Classes called `gridded components`, `coupler components` and `ESMF states` are used to achieve this functionality. The main objective of the ESMF state object is to enable a representation for data that needs to be shared between components. The gridded component class envelops a user component that accepts an import ESMF state and produces an export ESMF state. In a typical Earth system application using ESMF, land surface models, atmospheric models, ocean models etc. are defined as gridded components. The mapping of the import and export states, including spatial and temporal transformations, between different components are carried out by the coupler components. Typically user defined couplers are defined for each set of components that are coupled using ESMF.

To enable LIS to couple to other Earth system models using ESMF, the LIS code is enveloped using the ESMF superstructure layer. The motivation behind encompassing the LIS code with the ESMF superstructure is to enable LIS to act as the land modeling component in a coupled system. Further, since LSMs in LIS uses a common structure for data exchange (ALMA), each LSMs implemented in LIS can be used in a coupled application without having to adapt each LSM to be ESMF-compliant. The ESMF superstructure Figure 7 shows an example of a coupled application, where LIS is coupled to an ESMF compliant atmospheric model such as the Goddard Cumulus Ensemble (GCE [Tao et al., 2003]) model from GSFC. LIS acts as the gridded component, encapsulating all the included LSMs, such as CLM, Noah, and VIC. LIS and GCE interact through two custom defined couplers (LIS2GCE and GCE2LIS) that perform the mapping between ESMF states exchanged between the two components. The use of LIS helps not only in the use of different LSMs, but also in the use of high resolution data, high performance computing and data handling tools implemented in LIS.

5 Results

In this section we present some examples of the scientific studies and simulations enabled by LIS.

As described earlier, LIS's interoperable features enable a system with a growing suite of LSMs, forcing schemes, different types of parameter data, and domains. The extensible interfaces for these different components help in rapid prototyping and inclusion of new components into LIS. The configurable features in LIS also allow the user to select different components to build the application instead of having to use a monolithic system. For example, although LIS includes a number of land surface models and forcing schemes, the user can build the LIS application with the only required LSM and data scheme. These configurable features improve technology transfer and continued innovative efforts in modeling.

In the land surface modeling community, there have been numerous studies for performing intercomparisons between LSMs and sensitivity studies of specific parameterizations and forcings. Intercomparison studies such as the Global Soil Wetness Project (GSWP [Dirmeyer et al., 1999]) and Project for Intercomparison of Land-surface Parameterization Schemes (PILPS [Henderson-Sellers et al., 2003]) have provided significant insights to aid in future model and data set developments. One of the main challenges to performing such intercomparions is the configuration of each model for the specific set of data and domain. LIS provides an ideal platform to perform a multi-model land surface analysis and sensitivity studies of models to different parameters. As a demonstrative example, land surface simulations were carried out over Australia using three different models (Noah, CLM, and Mosaic). Sensible heat fluxes on October 9, 2001 at 4GMT predicted by the three models are shown in Figure 8. It can be observed that there are significant differences between the model predictions for the same day. Mosaic model in this case predicts higher values of sensible heat fluxes compared to that of CLM and Noah. Since the land surface model parameterizations vary significantly, an ensemble of LSMs is likely to represent the true distribution of responses in the Earth system. The suite of LSMs and common data make LIS an ideal platform to generate both model and ensemble predictions at different domains and resolutions.

One of the unique capabilities of LIS is the infrastructure it provides to support global land-atmosphere interactions at spatial resolutions down to 1km and finer. The high spatial resolution of LIS makes it capable of resolving spatial features such as urban areas that could not be resolved at coarser resolutions. Further, the ability to ingest high resolution data directly allows LIS to be a system that is capable of demonstrating the impact of high resolutions observations at the scale of observations themselves. Figures 9 and 10 show comparisons of the soils, topography, landcover, and LAI parameterizations at 1km and the coarser 1/4 degree resolution for an area around Fort Peck, MT. Fort Peck is part of two meteorological monitoring networks; the Surface Budget Radiation Network (SURFRAD [Augustine et al., 2000]) and the Coordinated Enhanced Observing Period (CEOP [Koike, 2004]). The dataset comparisons for Fort Peck are shown here as representative examples. Clearly, the 1km datasets capture more spatial heterogeneous features than the equivalent 1/4 degree datasets.

To compare and evaluate the predictions of the models, the model outputs (using CLM, Noah, and Mosaic) were compared with the surface observations at a number of locations (or reference sites) from the CEOP network. The model outputs at different resolutions were compared with the surface observations for a period from July 2001 to September 2001, also known as the first Enhanced Observing Period (EOP-1). In the comparisons, 5 different reference sites were used as shown in Table 2. As a representative example, the latent heat fluxes from CLM at the two resolutions were compared with the observations at Fort Peck, MT, and the resulting comparisons are shown in Figure 11 and 12. It can be seen that the 1km predictions are more improved than the 1/4 degree predictions which seem to have large positive biases. Similar results were obtained at other CEOP locations as well. A summary of the comparisons for different energy flux components, averaged over the CEOP reference sites and LSMs (represented as an ensemble) are shown in Figure 13. Clearly, the 1km predictions are more representative of the station data than the coarser domain predictions.

The high performance operating modes in LIS not only provides the ability to perform massive simulations, but also the environment to provide high throughput for the simulations. To demonstrate the improved performance with an increase in the number of processors, a global 1km simulation was performed on a custom built, 200 node Linux cluster at GSFC. The simulation was carried out using the Noah LSM and the GDAS forcing scheme. To manage the huge data throughput at 1km, the use of GDS servers to dynamically subset the data was used. Figure 14 shows the scaling of performance with the number of processors. The performance is compared with a theoretical $P/2$ estimate, which is based on the assumption that the performance improves by a factor of $P/2$ when P processors are employed. It can be seen that the LIS performance scales better than the $P/2$ estimate.

Uncoupled LSMs typically require many years of simulations to reach thermal and hydraulic equilibrium with the forcing meteorology. The adjustment process or the spin-up of the model that adjusts for initial anomalies in soil moisture content or meteorology is important in accurate characterization of the land surface conditions [Yang et al., 1995]. LIS include configurable features to dynamically subset datasets and configure simulations over the region of interest. These features enable in rapidly conducting spinups even at high spatial resolutions. For example, simulations for a year over a 1 deg x 1 deg domain at 1km spatial resolution with the Noah LSM requires approximately 1.2 hours using 4 processors. As a result, using 4 processors, LIS can perform 20 years of simulations in a day. The ability to spinup and generate initial conditions quickly is critically important for many applications using LIS.

The ability to use GDS servers for providing data allows LIS to perform simulations retrieving data from a data server that is located locally or remotely. This mode of operation helps in maintaining and managing a centralized, consistent datasets for multiple users. It also enables LIS be used in multiple applications, which can be run remotely by dynamically provisioning data in response to requests. The use of GDS servers helps in making datasets in varying formats transparent to the users. LIS currently uses a variety of data formats such as GRIB, NetCDF, and binary through the use of GDS server without necessarily implementing data access methods corresponding to each data standard.

The visualization tools in LIS allow for interactive use of both input and output data. The Land Explorer, which is based on the GDS server, allows user to interactively visualize and explore data at all resolutions. The LAS interface allows more advanced features such as performing data analysis, interactive subsetting etc. A snapshot of surface temperature for a region around Washington D.C. retrieved dynamically using the Land Explorer from a simulation using Noah model is shown in Figure 15.

6 Summary and Future Directions

LIS is an evolving framework for high resolution land surface modeling. The use of advanced software engineering concepts in the design of LIS provides a well defined architecture that allows the rapid specification of numerical models and data products. The interoperable features in LIS allow numerical models to explore various model/observation prediction scenarios for a given application.

In addition to providing a framework for land surface simulation, the flexible design of LIS enables researchers to focus on a wide variety of socially relevant science, education, application, and management issues. The accurate assessment of the spatial and temporal variation of the global land surface water can be used in conjunction with high resolution data obtained from satellites and other sources to improve our understanding of the natural processes. This knowledge can in turn be used for a more efficient management of natural resources. For example, the prediction of variables such as snowpack, amounts of soil moisture, the loss of water into the atmosphere from plants, etc. can be used to manage water in resource-limited areas. Researchers can also perform countless "what-if" studies, such as assessing the impact of landcover changes on climate change.

The process of modeling land surface globally at high resolutions is a *grand challenge* problem since it requires a significant resources in software, hardware, and communication performance. The use of scalable computing technologies in LIS enables previously computationally intractable problems to be handled in near real-time.

The interoperable framework provided by LIS provides parallel computing functions, data access and distribution, and interactive features. LIS software is implemented with the goals of high performance in near real-time, object-oriented design and interfacing allowing concurrent development of applications and software, ability to scale to different time/space resolutions, computer platforms, and land regions, and allow for easy incorporation of new model and data components. The interaction with other earth system models through ESMF allows LIS to participate in studies that investigate the nature of interaction and feedback between land and the atmosphere.

The flexibility and extensibility of LIS has enabled its use in numerous applications. As mentioned earlier, the incorporation of ESMF coupling structures allows LIS to interoperate with other ESMF-compliant systems. LIS currently uses the ESMF structures to couple with two different atmospheric models; the GCE model and the Weather Research and Forecasting (WRF [Michalakes et al., 2001]) model. The land modeling tools in LIS are designed to be the backbone of a prototype operational soil moisture modeling system used to aid the development of Future Combat Systems (FCS) for the US Army. A natural extension to LIS is the addition of carbon models to the water and energy dynamics to enable better understanding of the complex interactions of the physical domains that comprise the carbon cycle. Another key extension to the LIS infrastructure is the capability to perform data assimilation, which are techniques to merge observed data fields with model predictions to improve the subsequent predictions. LIS products can also be used in a number of other applications such as agricultural forecasting and water resources applications

LIS is evolving as a leading edge land surface modeling and data assimilation system to support broad land surface research and application activities. The use of the system by scientists, students, practitioners is expected to help in the effective application of high performance computing to high-resolution, real-time earth system studies and help in the development of earth system modeling and interoperability standards.

7 Acknowledgments

LIS is a Grand Challenge Investigation funded under NASA ESTO/CT CAN-00-OES-01 (Co-PI's Houser and Peters-Lidard), with additional support from NASA ESTO/AIST NRA-02-OES-04 (PI Peters-Lidard)

References

[ALMA, 2002] ALMA (2002). Assistance for land modeling activities, Version 3. http://www.lmd.jussieu.fr/ALMA/.

[Augustine et al., 2000] Augustine, J., DeLuisi, J., and Long, C. (2000). SURFRAD - A national surface radiation budget network for atmospheric research. *Bulletin of the American Meteorological Society*, 81(10):2341–2358.

[Dai et al., 2003] Dai, Y., Zeng, X., Dickinson, R., Baker, I., Bonan, G., Bosilovich, M., Denning, S., Dirmeyer, P., Houser, P., Niu, G., Oleson, K., Schlosser, A., and Yang, Z.-L. (2003). The common land model (CLM). *Bulletin of the American Meteorological Society*, 84(4):1013–1023. DOI:10.1175/BAMS-84-8-1013.

[Decyk et al., 1997] Decyk, V. K., Norton, C. D., and Szymanski, B. K. (1997). How to express C++ concepts in Fortran 90. *Scientific Programming*, 6(4):363–390.

[Dirmeyer et al., 1999] Dirmeyer, P., Dolman, A., and Sato, N. (1999). The global soil wetness project: A pilot project for global land surface modeling and validation. *Bulletin of the American Meteorological Society*, 80:851–878.

[Doty and Kinter, 1993] Doty, B. and Kinter, J. L. (1993). The grid analysis and display system (GrADS): A desktop tool for earth science visualization. In *American Geophysical Union 1993 Fall Meeting*, pages 6–10, San Francisco, CA.

[Ek et al., 2003] Ek, M., Mitchell, K., Yin, L., Rogers, P., Grunmann, P., Koren, V., Gayno, G., and Tarpley, J. (2003). Implementation of Noah land-surface model advances in the NCEP operational mesoscale Eta model. *Journal of Geophysical Research*, 108(D22). DOI:10.1029/2002JD003296.

[GTOPO30, 1996] GTOPO30 (1996). Global 30 arc second elevation data set. http://edcwww.cr.usgs.gov/landdaac/gtopo30/gtopo30.html.

[Hankin et al., 2001] Hankin, S. C., Callahan, J., and Sirott, J. (2001). The Live Access Server and DODS. In *17th International Conference on Interactive Information and Processing Systems for Meteorology, Oceanography, and Hydrology*, Albuquerque, NM.

[Hansen et al., 2000] Hansen, M., DeFries, R., Townshend, J., and Sohlberg, R. (2000). Global land cover classification at 1km spatial resolution using a classification tree approach. *International Journal of Remote Sensing*, 21(6):1331–1364.

[Henderson-Sellers et al., 2003] Henderson-Sellers, A., Irannejad, P., McGuffie, K., and Pitman, A. (2003). Predicting land-surface climates-better skill or moving targets? *Geophysical Research Letters*, 30(14):DOI:10.1029/2003GL017387.

[Hill et al., 2004] Hill, C., DeLuca, C., Balaji, V., Suarez, M., and da Silva, A. (2004). The architecture of the earth system modeling framework. *Computing in Science and Engineering*, 6(1).

[Kaufmann et al., 2000] Kaufmann, R., Zhou, L., Knyazikhin, Y., Shabanov, V., Myneni, R., and Tucker, C. (2000). Effect of orbital drift and sensor changes on the time series of AVHRR vegetation index data. *IEEE Transactions on Geoscience and Remote Sensing*, 38(6):2584–2597.

[Koike, 2004] Koike, T. (2004). The coordinated enhanced observing period - an initial step for integrated water cycle observation. *WMO Bulletin*, 53(2):115–121.

[Koster and Suarez, 1996] Koster, R. and Suarez, M. (1996). Energy and water balance calculations in the mosaic LSM. Technical memorandum 104606, NASA Goddard Space Flight Center.

[Liang et al., 1996] Liang, X., Lettenmaier, D., and Wood, E. (1996). One-dimensional statistical dynamic representation of subgrid spatial variability of precipitation in the two-layer variable infiltration capacity model. *Journal of Geophysical Research*, 101(D16):21403–21422.

[Liang et al., 1994] Liang, X., Lettenmaier, D., Wood, E., and Burges, S. (1994). A simple hydrologically based model of land surface water and energy fluxes for GCMs. *Journal of Geophysical Research*, 99(D7):14415–14428.

[Michalakes et al., 2001] Michalakes, J., Chen, S., Dudhia, J., Hart, L., Klemp, J., Middlecoff, J., and Skamarock, W. (2001). Development of a next generation regional weather research and forecast model. In Zwieflhofer, W. and Kreitz, N., editors, *Developments in Teracomputing: Proceedings of the Ninth ECMWF Workshop on the use of high performance computing in meteorology*, pages 269–276, Singapore.

[Miller and White, 1998] Miller, D. and White, R. (1998). A conterminous United States multi-layer soil characteristics data set for regional climate and hydrology modeling. *Earth Interactions*, paper 1.

[Mitchell et al., 2004] Mitchell, K. E., Lohmann, D., Houser, P. R., Wood, E. F. Schaake, J. C., Robock, A., Cosgrove, B., Sheffield, J., Duan, Q., Luo, L., Higgins, W. R., Pinker, R. T., Tarpley, J. D., Lettenmaier, D. P., Marshall, C. H., Entin, J. K., Pan, M., Shi, W., Koren, V., Meng, J., Ramsay, B. H., and Bailey,

A. A. (2004). The Multi-institution North American Land Data Assimilation system (NLDAS): Utilization of multiple GCIP products and partners in a continental distributed hydrological modeling system. *Journal of Geophysical Research*, 109:DOI:10.1029/2003JD003823.

[Nowack, 1997] Nowack, P. (1997). Architectural abstractions for frameworks. In Bosch, J. and Mitchell, S., editors, *Lecture Notes in Computer Science, Object-Oriented Technology, ECOOP'97 Workshops*, volume 1357, pages 116–118, Jyvaskyla, Finland. Springer Verlag.

[Pree, 1995] Pree, W. (1995). *Design Patterns for Object-Oriented Software Development*. Addison-Wesley.

[Reynolds et al., 1999] Reynolds, C., Jackson, T., and Rawls, W. (1999). Estimating available water content by linking the FAO soil map of the world with global soil profile database and pedo-transfer functions. In *American Geophysical Union Fall Meeting, EOS Transactions*, page 80.

[Rodell et al., 2004] Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng., C.-J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J. K., Walker, J. P., Lohmann, D., and Toll, D. (2004). The Global Land Data Assimilation System. *Bulletin of the American Meteorological Society*, 85(3):381–394.

[Sellers et al., 1986] Sellers, P. J., Mintz, Y., and Dalcher, A. (1986). A simple biosphere model (SiB) for use within general circulation models. *Journal of Atmospheric Science*, 43:505–531.

[Sud and Mocko, 1999] Sud, Y. and Mocko, D. (1999). New snow-physics to complement SSiB. part I: Design and evaluation with ISLSCP initiative I datasets. *Journal of Meteorological Society of Japan*, 77(1B):335–348.

[Tao et al., 2003] Tao, W.-K., Simpson, J., Baker, D., Braun, S., Chou, M.-D., Ferrier, B., Johnson, D., Khain, A., Lang, S., Lynn, B., Shie, C.-L., Starr, D., Sui, C.-H., Wang, Y., and Wetzel, P. (2003). Microphysics, radiation and surface processes in the Goddard Cumulus Ensemble (GCE) model. *Meteorology and Atmospheric Physics*, 82:97–137.

[Tian et al., 2002] Tian, Y. Woodcock, C., Wang, Y., Privette, J., Shabanov, N., Zhou, L., Zhang, Y., Buermann, W., Dong, J., Veikkanen, B., Hame, T., Andersson, K., Ozdogan, M., Knyazikhin, Y., and Myneni, R. (2002). Multiscale analysis and validation of the MODIS LAI product I. uncertainty assessment. *Remote Sensing of Environment*, 83:414–430.

[Wielgosz et al., 2001] Wielgosz, J., Doty, B., Gallagher, J., and Holloway, D. (2001). GrADS and DODS. In *Seventeenth International Conference on Interactive Information and Processing*, Albuquerque, NM.

[Yang et al., 1995] Yang, Z.-L., Dickinson, R.E. Henderson-Sellers, A., and Pitman, A. (1995). Preliminary study of spin-up processes in land surface models with the first stage of project for intercomparison of land surface parameterization scheme phase 1(a). *Journal of Geophysical Research*, 100(16):553–578.

Table 1
Summary of user defined components implemented in LIS

Type	Name	Source
Land Surface	Community Land Model (CLM) [Dai et al., 2003]	NCAR
Model	Noah [Ek et al., 2003]	(NCEP)
	Variable Infiltration Capacity (VIC)	UW and PU
	[Liang et al., 1994,Liang et al., 1996]	
	Mosaic [Koster and Suarez, 1996]	GSFC
	Hydrology with Simple SIB (HySSIB)	GSFC
	[Sellers et al., 1986,Sud and Mocko, 1999]	
Forcing	Goddard Earth Observing System (GEOS)	GSFC
Data	Global Data Assimilation System (GDAS)	NCEP
	European Center for Medium Weather	ECMWF
	Forecasting (ECMWF)	
	North American Land Data Assimilation System (NLDAS)	GSFC
	Agricultural Meteorological Modeling System (AGRMET)	AFWA
	CPC's Merged Analysis of precipitation (CMAP)	CPC
	CPC's Morphing Technique (CMORPH)	CPC
	precipitation product	
Parameter	Global 1km static landcover map [Hansen et al., 2000]	UM
Data	Global 1km topography (GTOPO30) [GTOPO30, 1996]	USGS
	Soils data from State Soil Geographic	NRCS
	Database (STATSGO) [Miller and White, 1998]	
	Soils data from [Reynolds et al., 1999]	NOAA
	AVHRR Leaf Area Index (LAI) [Kaufmann et al., 2000]	BU
	MODIS LAI [Tian et al., 2002]	BU

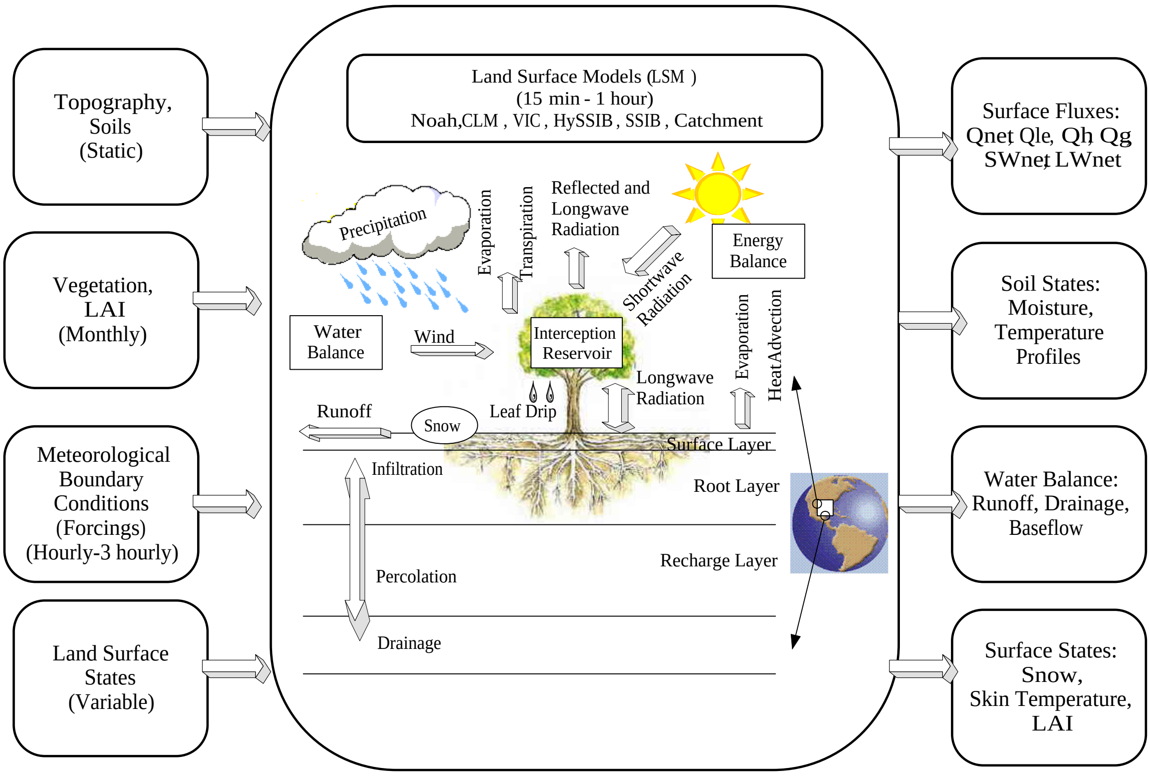

Fig. 1. A schematic representation of land surface modeling in LIS

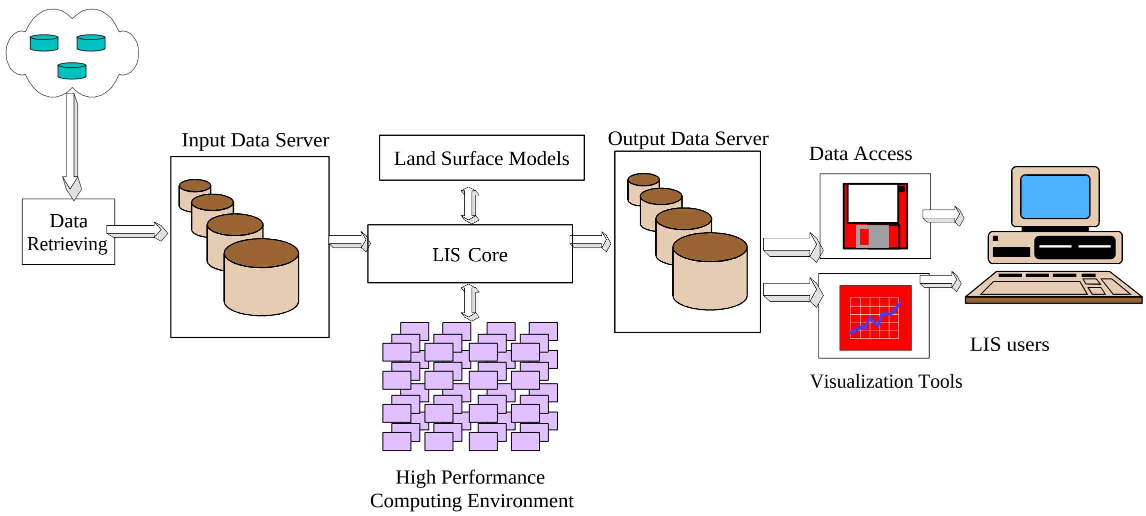

Fig. 2. Components of the LIS software architecture

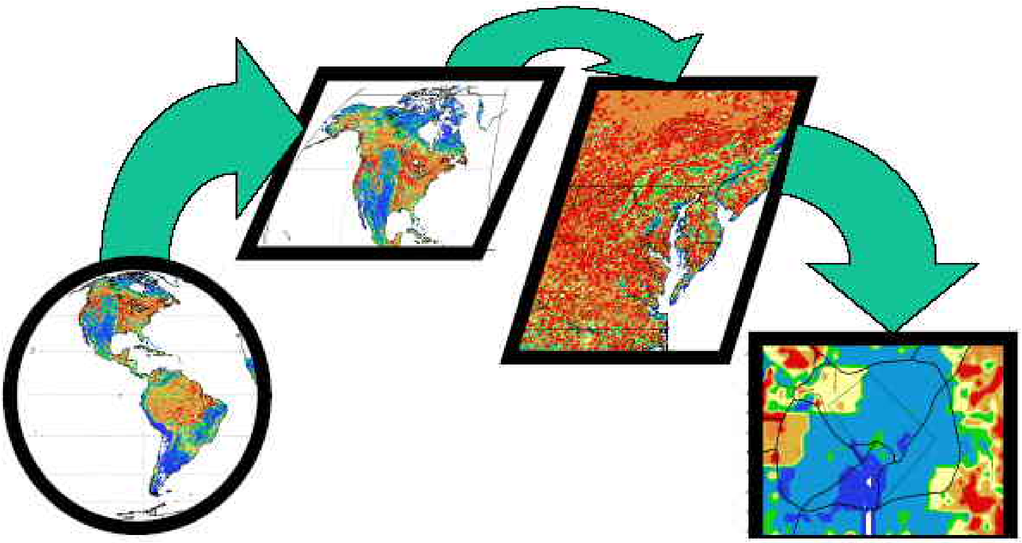

Fig. 3. 1km resolution global MODIS Leaf Area Index product

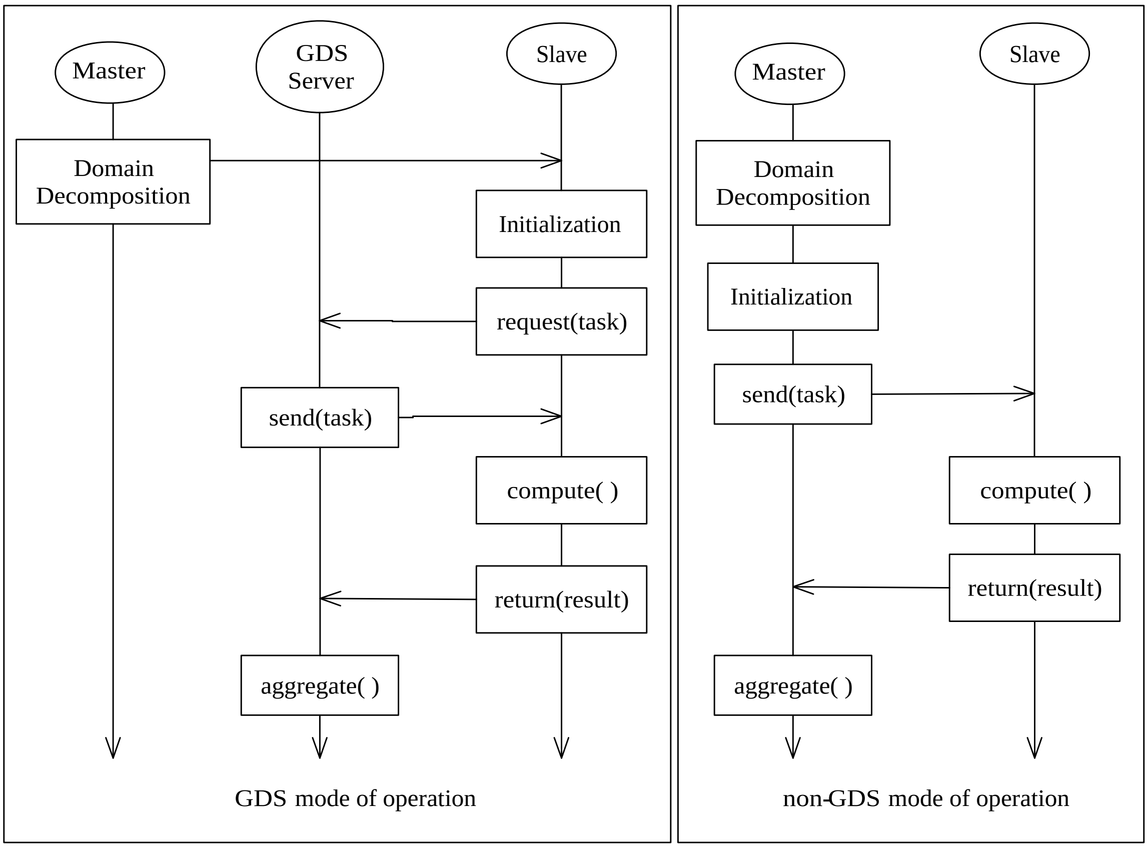

Fig. 4. Roles of master, slave, and the GDS server in different parallel modes of operation in LIS

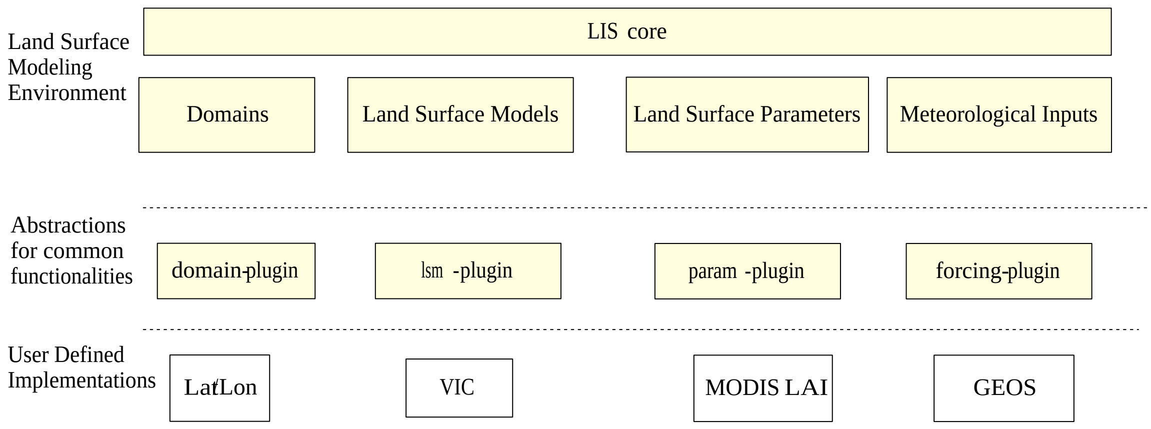

Fig. 5. Layered architecture of the LIS components

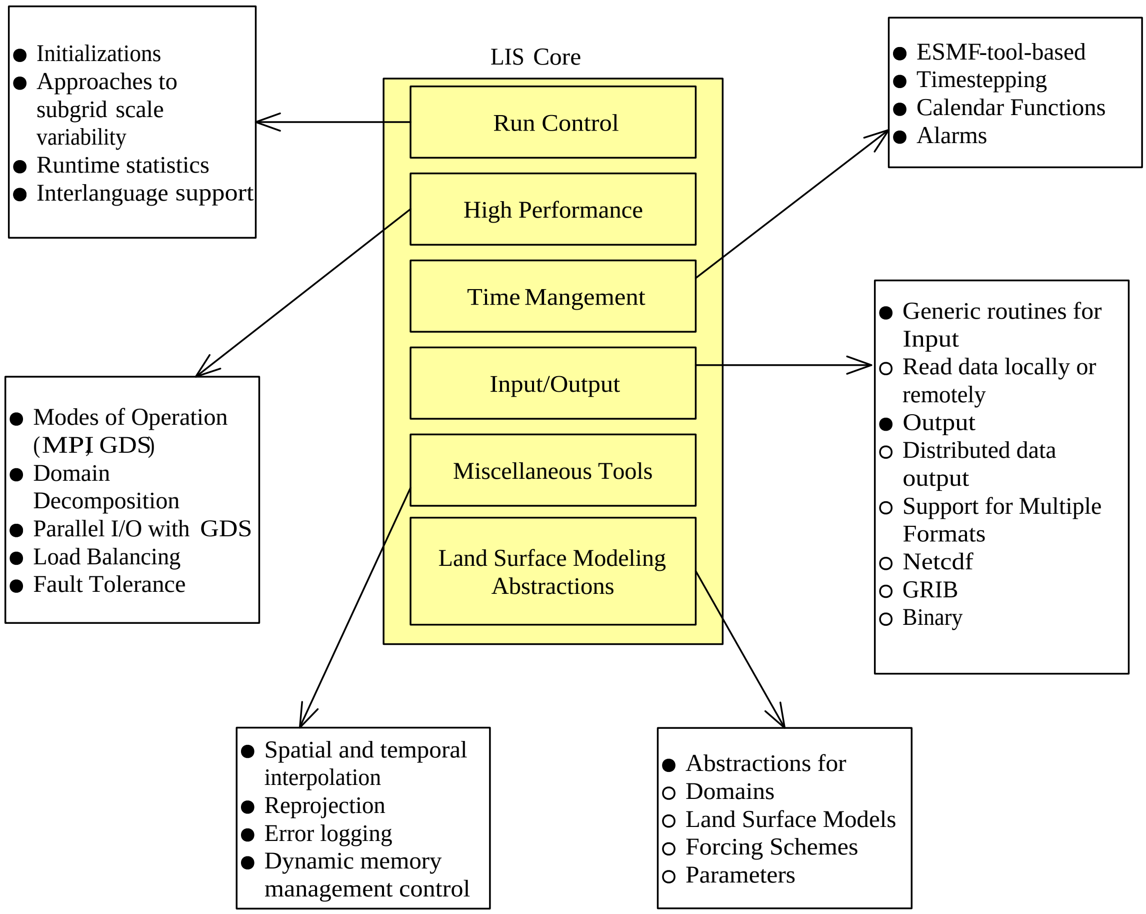

Fig. 6. Main functions of the LIS core

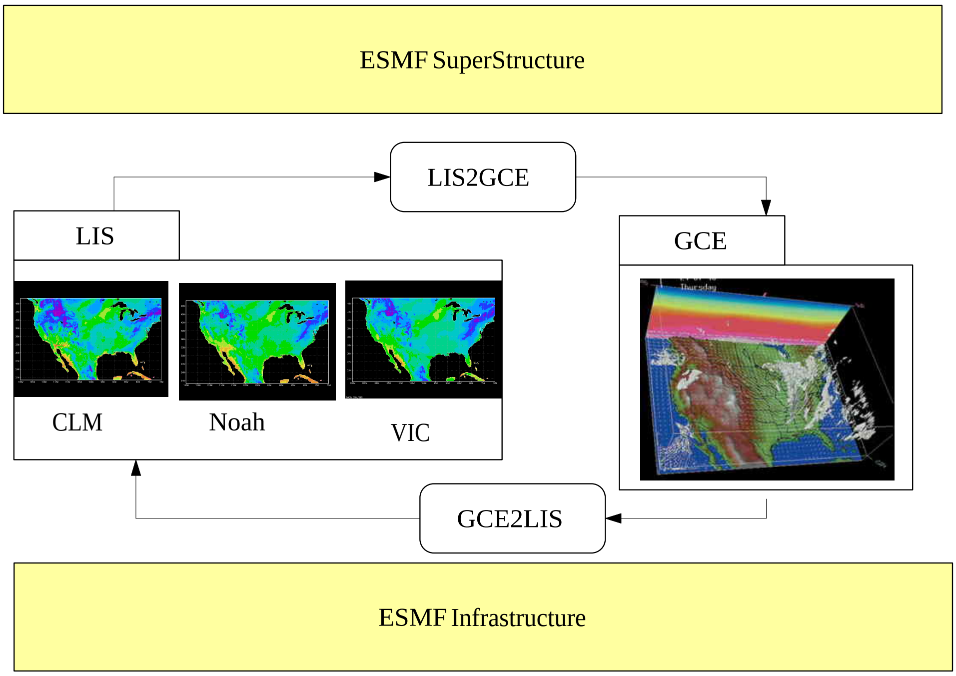

Fig. 7. A simple example of an application where the LSMs in LIS are coupled to GCE using ESMF

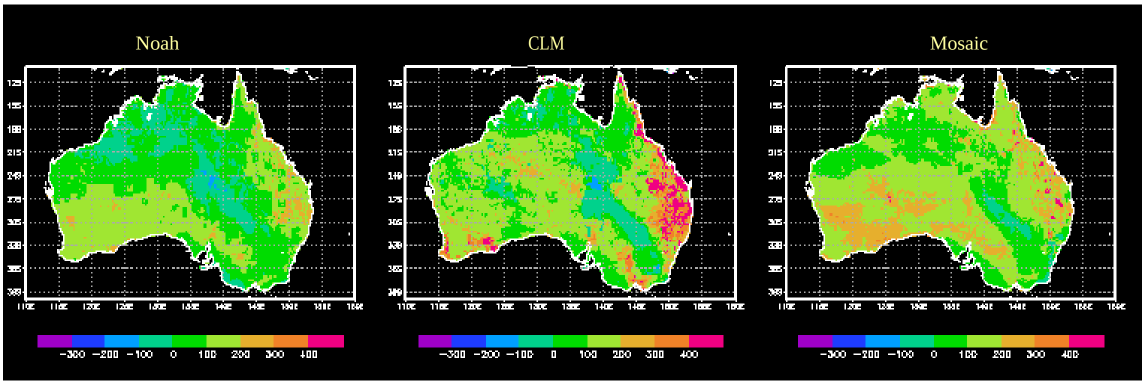

Fig. 8. Sensible heat fluxes (W/m2) produced by Noah, CLM, and Mosaic over Australia, at 4GMT for October 09, 2001.

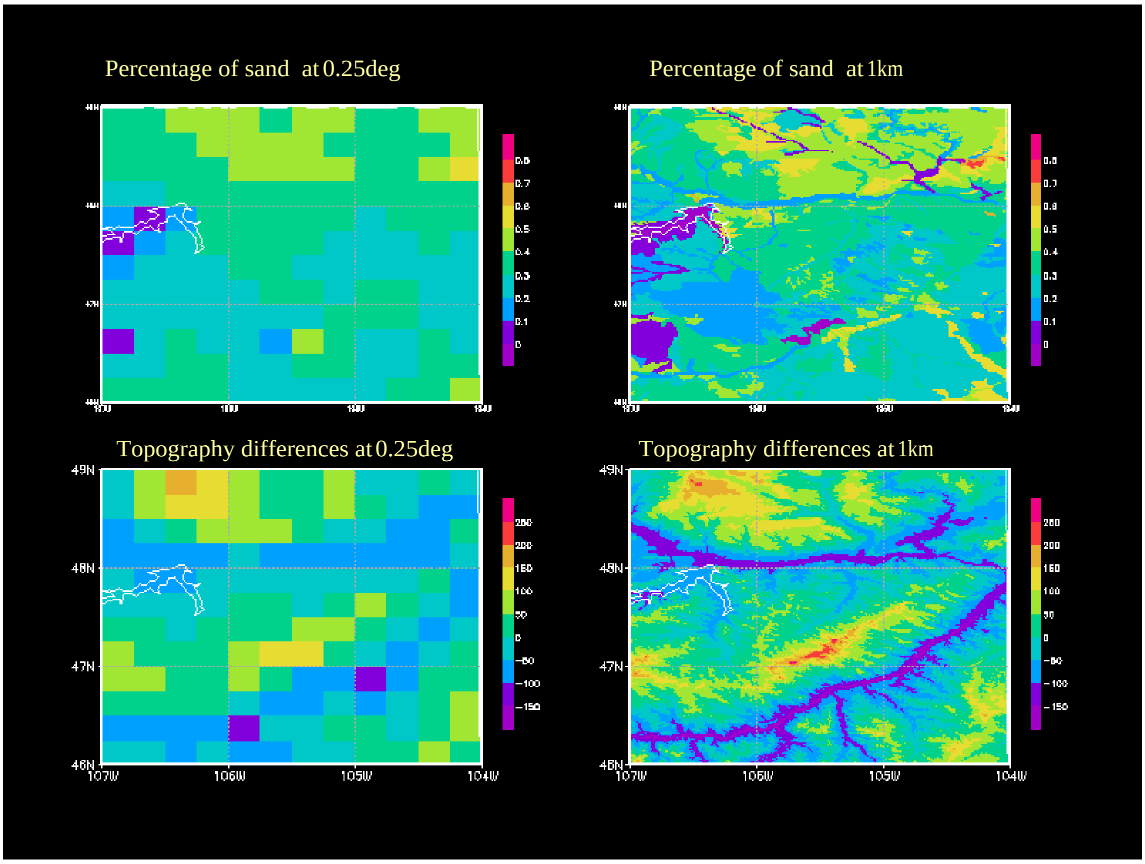

Fig. 9. Comparison of the 1km and 1/4 degree data sets for Soil (sand fraction), and Topography differences (between the 1km topography and model) for Fort Peck, MT.

Table 2
CEOP reference sites used for intercomparisons with LIS model outputs

Location	Latitude	Longitude
Fort Peck, MT	48.31N	105.10W
Bondville, IL	40.01 N	88.29W
Cabauw, Netherlands	51.97N	4.927E
Mongolia	45.74-46.78N	106.26-107.48E
Brasilia, Brazil	15.93S	47.92W

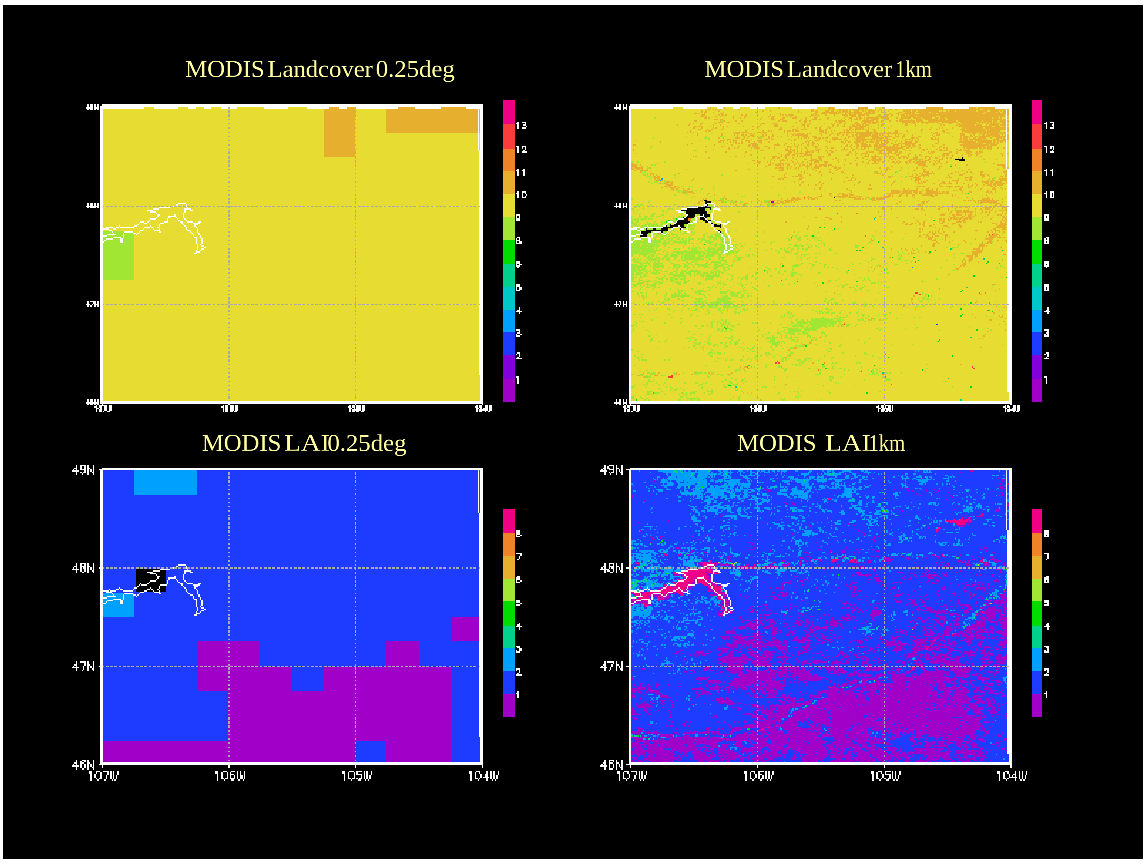

Fig. 10. Comparison of the 1km and 1/4 degree data sets for landcover and Leaf Area Index (LAI) from MODIS datasets for Fort Peck, MT.

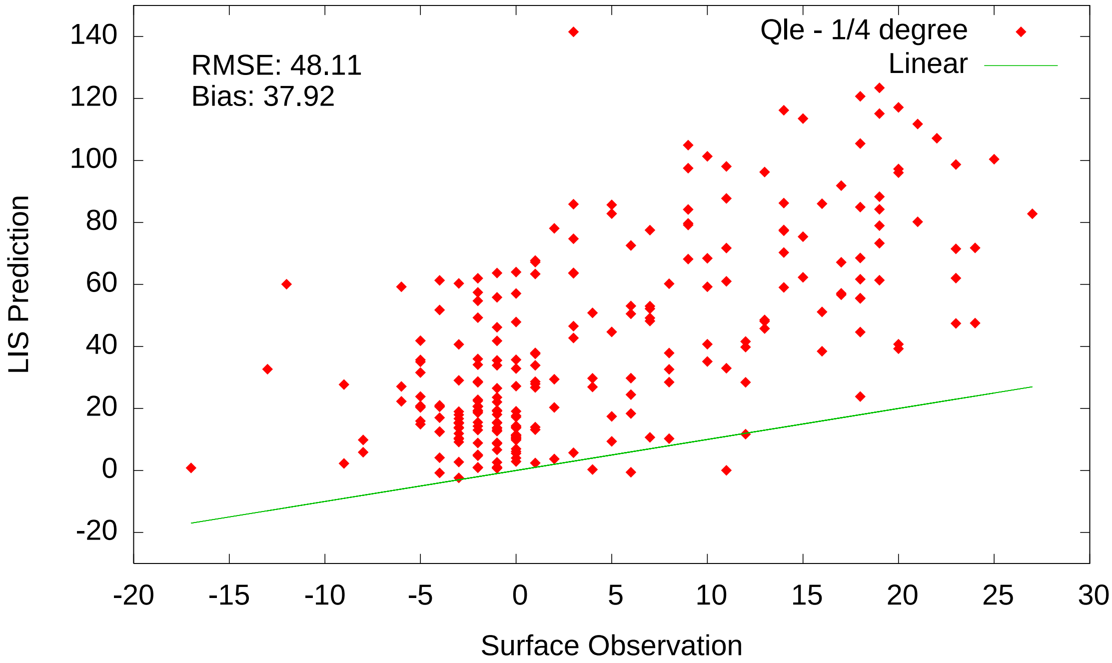

Fig. 11. Comparison of the Latent Heat Fluxes from CLM at 1/4 degree resolution with the Surface Observations at Fort Peck, MT. for the EOP-1 period.

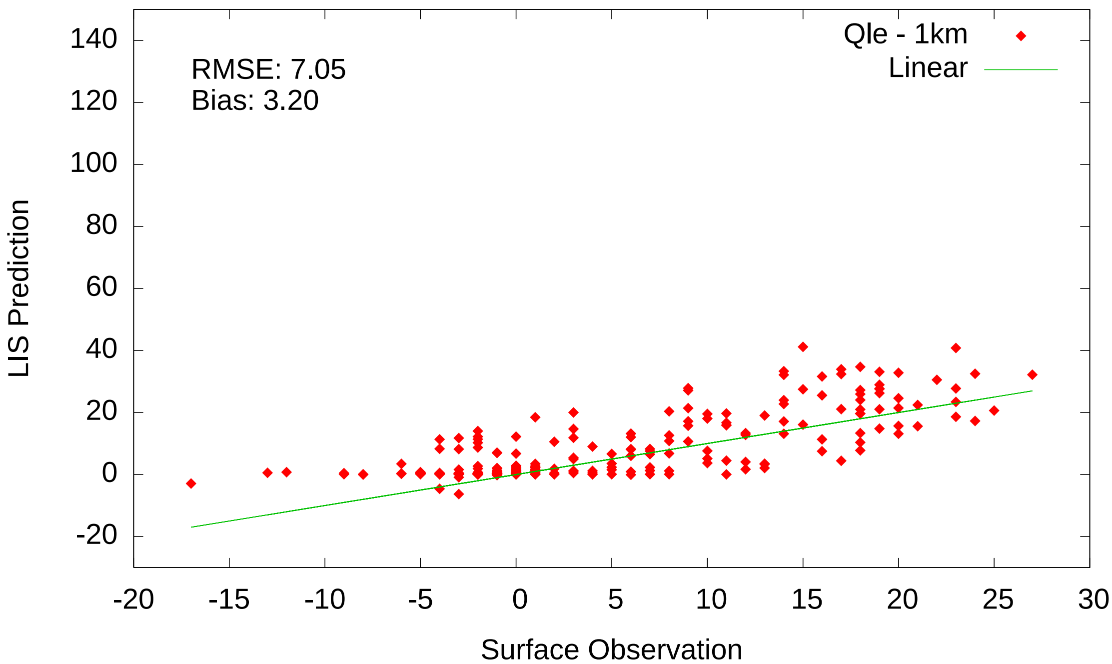

Fig. 12. Comparison of the Latent Heat Fluxes from CLM at 1km resolution with the Surface Observations at Fort Peck, MT. for the EOP-1 period.

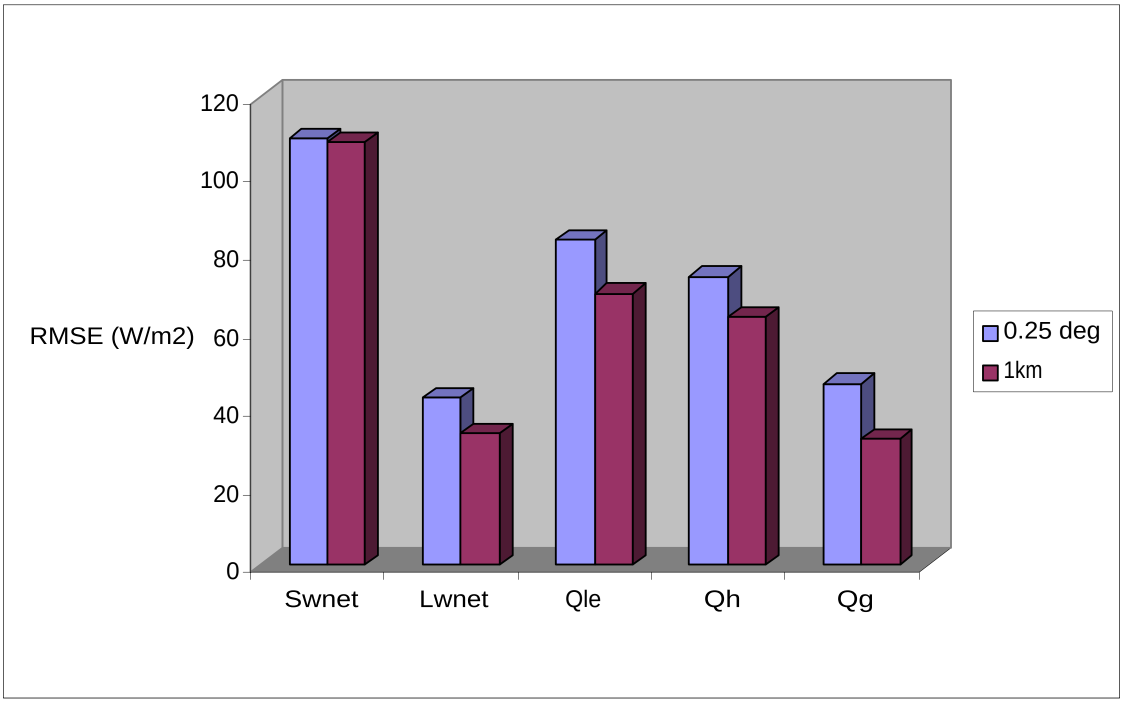

Fig. 13. Comparison of Root Mean Square Errors (RMSE) at 1/4 degree and 1km resolution for the energy fluxes averaged over all LSMs and CEOP sites ($SWnet$: Net shortwave Radiation, $LWnet$: Net longwave radiation, Qle - Latent heat flux, Qh - Sensible heat flux, Qg - Ground heat flux)

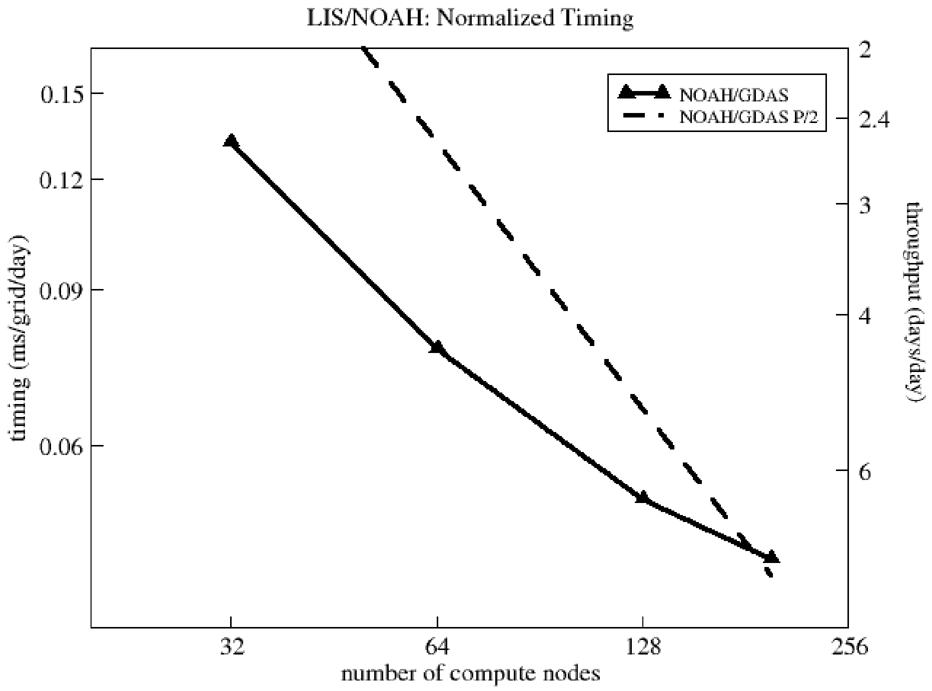

Fig. 14. Normalized timing and simulation throughput for a global 1km LIS simulation with Noah using the GDAS forcing compared with the theoretical $P/2$ estimate.

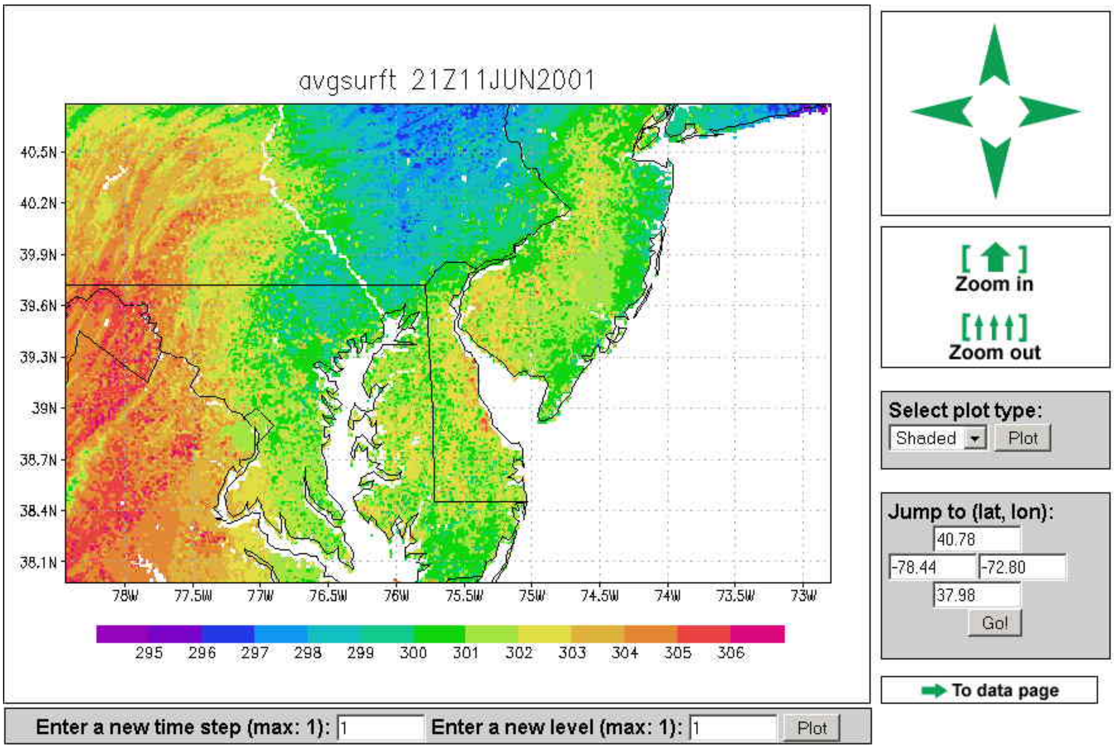

Fig. 15. A snapshot of the Surface Temperature datafield for a region around Washington D.C. retrieved dynamically by the Land Explorer from a global 1km output of a Noah LSM simulation